10 March 1970

MEMORANDUM FOR: Mr. John K. Vance

SUBJECT: Subject-Keyword Linkage

     I have examined the two sample printouts demonstrating linkages
between subject codes and keywords. Although the subject-keyword
listing is disappointing in that many associations occur only once (and
thus may be spurious), nevertheless the printouts are most interesting.
I still feel that a complete printout of associations (using a cutoff to avoid
printing linkages that do not occur at least X times) would be very useful
both as a search tool and as a core of relationships to use in the compilation
of a controlled keyword list. Indeed it is possible that a complete printout
of significant associations could function as a controlled vocabulary for use
in searching with very little additional work being necessary.

     Jointly the two complementary printouts do exactly what I expected
them to do. The subject-keyword list is an aid in the construction of
search strategies in the situation in which one wants to use keywords to
obtain greater specificity. Thus, if we want to search on commerical
aviation activities in a particular country, we need to search on area
code, subject code (430.000) and those specific keywords that indicate
commercial aviation. An examination of the keyword list under 430.000
would tell us that we need to search on CIVIL, other combinations
beginning with CIVIL, COMMERCIAL, and a list of keywords indicating
specific carriers (e.g., TWA).

     The keyword-subject listing can be used to indicate the context in
which a particular keyword has been used. Suppose we want to search
on labor strikes in the Middle East. We want to avoid retrieval of items
on military strikes and strike forces. An examination of the printout
under the keyword STRIKE indicates that it has been used in the context

of 220.000 and 112.200, both of which have labor connotations. However, it has also been used with 122.200 (anti-government activities) and 209.900 (public order and safety) and in these contexts it may have been used with a different connotation. Going further down the list we find the association of STRIKE with 128.000 (insurgency) which presumably indicates a military strike rather than a labor strike, with 206.000 (education) which presumably indicates a strike of teachers, with 123.000 (internal security) which is more likely to indicate the military version, or with 452.000 (electronic communications systems); the relationship of the latter code to STRIKE is not immediately apparent (to me at least). In other words, if we want to conduct a comprehensive search on the occurrence of strikes (labor) in, say, the Arab countries, the listings under STRIKE and STRIKES will indicate which subject codes we need to search to ensure maximum retrieval without too much irrelevancy.

Incidentally some keyword assignment appears to be largely redundant. The keyword PUBLIC SAFETY, for example, occurs 287 times with the code 209.900 (public order and safety) which is surely a redundancy. The use of specific public safety keywords (e.g., FIRE PROTECTION) with 209.900 would be necessary. Likewise, the use of the keyword PUBLIC SAFETY with other subject codes would be significant. But there is little point in organizing a general keyword that is virtually co-extensive with the subject code with which it is used.

25X1A9a